Session 10

Assignment 1

Task 1:

1. **Blood glucose levels for obese patients have a mean of 100 with a standard deviation of 15. A researcher thinks that a diet high in raw cornstarch will have a positive effect on blood glucose levels. A sample of 36 patients who have tried the raw cornstarch diet have a mean glucose level of 108. Test the hypothesis that the raw cornstarch had an effect or not.**

   SOLUTION:

   Step 1: State the null hypothesis: H0: μ=100

   Step 2: State the alternate hypothesis: H1: $\neq$100

   Step 3: State the alpha level. For this instance, use 0.05 as alpha level. As this is a two-tailed test, split the alpha into two.

   0.05/2=0.025

   Step 4: Find the z-score associated with the alpha level. Look for the area in one tail only. A z-score for 0.75(1-0.025 = 0.975) is 1.96. As this is a two-tailed test, also consider the left tail (z = 1.96)

   Step 5: Find the test statistic using the following formula.

   $$Z = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$$

   z = (108-100) / (15/√36) = 3.2

   Step 6: If Step 5 is less than -1.96 or greater than 1.96 (Step 3), reject the null hypothesis.

   In this case, it is greater, so **reject the null.**

---

2. **In one state, 52% of the voters are Republicans, and 48% are Democrats. In a second state, 47% of the voters are Republicans, and 53% are Democrats. Suppose a simple random sample of 100 voters are surveyed from each state.**

   What is the probability that the survey will show a greater percentage of Republican voters in the second state than in the first state?

   <u>SOLUTION:</u>

- Let P1 = the proportion of Republican voters in the first state, P2 = the proportion of Republican voters in the second state, p1 = the proportion of Republican voters in the sample from the first state, p2 = the proportion of Republican voters in the sample from the second state.
- The number of voters sampled from the first state (n1) = 100 and the number of voters sampled from the second state (n2) = 100.
- Check if the sample size is big enough to model differences with a normal population. n1*P1 = 100 * 0.52 = 52, n1*(1 - P1) = 100 * 0.48 = 48, n2*P2 = 100 * 0.47 = 47, n2*(1 - P2) = 100 * 0.53 = 53 are each greater than 10. Thus, the sample size is large.
- Find the mean of the difference in sample proportions:
  E(p1 - p2) = P1 - P2 = 0.52 - 0.47 = 0.05.
- Find the standard deviation of the difference.
  σd = sqrt{ [P1(1 - P1) / n1] + [P2(1 - P2) / n2] }
  σd = sqrt{ [(0.52)(0.48) / 100] + [(0.47)(0.53) / 100] }
  σd = sqrt (0.002496 + 0.002491) = sqrt(0.004987) = 0.0706
- Find the probability. This problem requires us to find the probability that p1 is less than p2. This is equivalent to finding the probability that p1 - p2 is less than zero. To find this probability, we need to transform the random variable (p1 - p2) into a z-score. That transformation appears below.
  z p1 - p2 = (x - μ p1 - p2) / σd = = (0 - 0.05)/0.0706 = -0.7082
- Using Normal Distribution Calculator, we find that the probability of a z-score being -0.7082 or less is 0.24.
- Therefore, the probability that the survey will show a greater percentage of Republican voters in the second state than in the first state is **0.24**.

3. **You take the SAT and score 1100. The mean score for the SAT is 1026 and the standard deviation is 209. How well did you score on the test compared to the average test taker?**

SOLUTION:

Step1: Write the X-value into the z-score equation. For this sample, the X-value is the SAT score i.e. 1100.

$$Z = \frac{1100 - \mu}{\sigma}$$

Step 2: Substitute the mean μ into the z-score equation.

$$Z = \frac{1100 - 1026}{\sigma}$$

Step 3: Write the standard deviation σ into the z-score equation.

$$Z = \frac{1100 - 1026}{209}$$

Step 4: Calculate the answer as shown below:

(1100 – 1026) / 209 = 0.354

This means that the score was 0.354 standard deviation above the mean.

Step 5: Look up the z-value in the z-table to see what percentage of test-takers scored below the score.

A z-score of 0.354 = **0.6368** or **63.68%.**

Task 2:

1. **Is gender independent of education level? A random sample of 395 people were surveyed and each person was asked to report the highest education level they obtained. The data that resulted from the survey is summarized in the following table:**

	High School	**Bachelors**	**Masters**	**Ph.D.**	**Total**
Female	**60**	**54**	**46**	**41**	**201**
Male	**40**	**44**	**53**	**57**	**194**
Total	**100**	**98**	**99**	**98**	**395**

Question: Are gender and education level dependent at 5% level of significance? In other words, given the data collected above, is there a relationship between the gender of an individual and the level of education that they have obtained?

SOLUTION:

Below is the table of expected counts based on the formula as shown.
E = (row total × column total) / sample size

E	High School	Bachelors	Masters	Ph.D.	Total
Female	50.886	49.868	50.377	49.868	201
Male	49.114	48.132	48.623	48.132	194
	100	98	99	98	395

Chi-Square Test Statistic

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

=(60−50.886)2/50.886+…+(57−48.132)2/48.132 = **8.006**

The critical value of **χ2** with 3 degree of freedom = **7.815**

Since 8.006 > 7.815 therefore, **reject the null hypothesis** and conclude that the education level depends on gender at a 5% level of significance.

2. **Using the following data, perform a oneway analysis of variance using α=.05. Write up the results in APA format.**

   [Group1: 51, 45, 33, 45, 67]
   [Group2: 23, 43, 23, 43, 45]
   [Group3: 56, 76, 74, 87, 56]

   <u>SOLUTION:</u>

The sample mean ($\bar{X}$) for Group1 = 48.2, Group2 = 35.4, Group3 = 69.8

Intermediate steps in calculating the Group Variances is given below.

Group1 Values	Mean	Deviations	Sq. Deviations
51	48.2	2.8	7.84
45	48.2	-3.2	10.24
33	48.2	-15.2	231.04
45	48.2	-3.2	10.24
67	48.2	18.8	353.44

Group2 Values	Mean	Deviations	Sq. Deviations
23	35.4	-12.4	153.76
43	35.4	7.6	57.76
23	35.4	-12.4	153.76
43	35.4	7.6	57.76
45	35.4	9.6	92.16

Group3 Values	Mean	Deviations	Sq. Deviations
56	69.8	-13.8	190.44
76	69.8	6.2	38.44
74	69.8	4.2	17.64
87	69.8	17.2	295.84
56	69.8	-13.8	190.44

Sum of squared deviations from the mean (SS) for the groups are as follows.
For Group1 = 612.8, Group2 = 515.2, Group3 = 732.8

Var1 = 612.8 / (5−1) = 153.2
Var2 = 515.2 / (5−1) = 128.8
Var3 = 732.8 / (5−1) = 183.2

MSerror = (153.2+128.8+183.2) / 3 = 155.07
Note: this is just the average within-group variance; it is not sensitive to group mean differences.

Calculating the remaining error (or within) terms for the ANOVA table:
dferror = 15−3 = 12
SSerror = (155.07)(15−3) = 1860.8
Intermediate steps in calculating the variance of the sample means are as follows.

Grand Mean ($\bar{X}_{grand}$) = (48.2+35.4+69.8) / 3 = 51.13

Group Mean	Grand Mean	Deviations	Sq. Deviations
48.2	51.13	-2.93	8.58
35.4	51.13	-15.73	247.43
69.8	51.13	18.67	348.57

Sum of squares (SSmeans) = 604.58

Varmeans = 604.58 / (3−1) = 302.29

MSbetween = (302.29)(5) = 1511.45

Calculating the remaining between (or group) terms of the ANOVA table:

dfgroups = 3−1 = 2

SSgroup = (1511.45)(3−1) = 3022.9

Test statistic and critical value

F = 1511.45 / 155.07 = 9.75

Fcritical(2,12) = 3.89

Decision: reject H0

ANOVA table

source	SS	df	MS	F
group	3022.9	2	1511.45	9.75
error	1860.8	12	155.07	
total	4883.7			

Effect size

η^2=3022.9/4883.7=0.62

APA writeup :

$F(2, 12)=9.75$, $p < 0.05$, η^2

= 0.62

3. **Calculate F Test for given 10, 20, 30, 40, 50 and 5,10,15, 20, 25.**

For 10, 20, 30, 40, 50:

<u>SOLUTION:</u>

Calculate Variance of first set

Total Inputs (N) =(10,20,30,40,50)

Total Inputs (N)=5

Mean (xm)= (x1+x1+x2...xn)/N

Mean (xm)= 150/5

Means(xm)= 30

SD=sqrt(1/(N-1)*((x1-xm)2+(x2-xm)2+..+(xn-xm)2))

=sqrt(1/(5-1)((10-30)2+(20-30)2+(30-30)2+(40-30)2+(50-30)2))

=sqrt(1/4((-20)2+(-10)2+(0)2+(10)2+(20)2))

=sqrt(1/4((400)+(100)+(0)+(100)+(400)))

=sqrt(250)

=15.8114

Variance=SD2

Variance=15.81142

Variance=250

Calculate Variance of second set

For 5, 10,15,20,25:

Total Inputs(N) =(5,10,15,20,25)

Total Inputs(N)=5

Mean (xm)= (x1+x2+x3...xN)/N

Mean (xm)= 75/5

Means (xm)= 15

SD=sqrt(1/(N-1)*((x1-xm)2+(x2-xm)2+..+(xn-xm)2))

=sqrt(1/(5-1)((5-15)2+(10-15)2+(15-15)2+(20-15)2+(25-15)2))

=sqrt(1/4((-10)2+(-5)2+(0)2+(5)2+(10)2))

=sqrt(1/4((100)+(25)+(0)+(25)+(100)))

=sqrt(62.5)

=7.9057

Variance=SD2

Variance=7.90572

Variance=62.5

To calculate F Test

F Test = (variance of 10, 20,30,40,50) / (variance of 5, 10, 15, 20, 25)

= 250/62.5

= 4.

The F Test value is 4.